A

Synopsis

On

A Novel approach for the database Creation and validation using Master Data Management

ए नावेल अप्प्रोच फॉर दडेटाबेस क्रिएशनएंडवेलिडेशन युसिंग मास्टर डाटा मैनेजमेंट

For Registration

Of

DOCTOR OF PHILOSOPHY

In

Computer Science and Applications

Submitted By

Dupinder Kaur

Research Scholar

Supervised By

Dr. Dilbag Singh,Professor

Dept. of Computer Science & Applications

Department of Computer Science and Applications

Chaudhary Devi Lal University, SIRSA-125055

(2017-18)

Contents

1. Introduction

1.1 Master Data Management: Overview

1.2 Process of Master Data Creation

1.3 Master Data Governance

2. Research Methodology

2.1 Problem Statement

2.2 Objectives of the Study

2.3 Scope of the Study

2.4 Research Methodology Used in Research

2.5 Chapterization

3. Literature Review

4. References

1 Introduction

In today's rapidly changing business environment, the emphasis on enterprise lightsomeness is changing into progressively vital wherever business executive, square measures trying to become faster and improved deciding by harnessing the information hold on within the enterprise by means of that knowledge and therefore the quality of storing the info in pregnant ways in which, that corporations will later take advantage (e.g. mistreatment analytical systems).As enterprises grows, the quantity and complexity of storing the data becomes difficult. This complexity has created a number of challenges in the ways that many enterprises has previously managed their data and is especially important to addresses most critical data entities(golden record) also known as *master data*. Master data management (MDM) is an umbrella term for a set of technology independent disciplines focusing on master data with the common goal of helping enterprises to achieve a single version of the truth for its master entities, which in today's enterprises are usually spread across heterogeneous data sources which has historically made it difficult for some enterprise to achieve a collective view on, for example, who their customers are and if they are really meeting the market demands [5].

Master data is the complete record of a human like the data is about customers, patients, suppliers, partners, employees, and other critical entities. It is the basis for the high-value, core information that is used to support critical business processes across the enterprise. MDM is the heart of every transaction, application, report and decision of an enterprise. Organizations may hold and replicate this master data across many different applications and customer touch points, such as order processing, customer service, and reporting systems. However, there can be the problem of inconsistency of data because the way how the data is created, updated, and maintained is done by their own unique way. The entered data may be of low quality because of some missing, incomplete, duplicated entries. Critical business processes are hampered by incorrect data that can lead to poor decisions and result in poor business outcomes [6].

According to the existing literature and experiences in data cleaning project, business and data quality rules, data controlling, and data cleaning project planning and implementation are described in order to maintain data integrity in the case company. In addition, it is proven that data cleaning actions can be done manually, making it more affordable to the case company than acquiring an external data cleaning service. The results of this study may be used in future

cleaning projects inside and outside the company. [4] In the future it would be interesting to study what is the most favorable combination of manual and external system based cleaning for the case company [10].

1.1 Master data management (MDM): Overview

The main objective of MDM is to provide authoritative master data to an enterprise known as a *system of record* (SOR). The SOR is the one source where master data is guaranteed to be accurate, up-to-date and thus the best source of truth.[3]

According to David Loshin master data management (MDM) comprises a set of data management best practices to help corporations' key stakeholders and business clients with incorporating business applications, information management methods, and data management tools in an endeavor to improve and maintain clean, accurate, timely and complete data across disparate applications.[2]

For effective result of database MDM consist of 6 building blocks which are necessary for result.

1. **Vision**: The scope of the MDM can be described, for example the domains which it is affecting. It needs to be clearly stated how the MDM vision supports the company's business vision. The business justification for the project has to be clear and enduring. Similar to business vision, there can only be one MDM vision. As sustainability and responsibility are part of the case company's business vision, the vision of the MDM project described in this research is aligned to that vision, as it is to ensure high level in quality, accuracy, availability and completeness of master data.[7]

2. **Strategy:** The need of how MDM vision will be realized and how to manage the master data assets within the organization is important assets. There is a requirement of data owner and manner in which data will be enriched.
Where the data will be stored and where it will be published?
Who needs to consume it and who will have the rights to modify it?

These requirements need to be analyzed to determine the priorities. The result will be a prioritized road map; an implementation plan. The case company's strategy is to reach the MDM objectives by analyzing the current situation in detail for five domains and based on that the future governance, processes, data models and business rules for managing the master data will be designed. [7]

3. **Governance**: Without effective governance, an MDM initiative will probably fail. Gartner defines MDM governance as the specification of decision rights and accountability framework to encourage desirable behavior in the ongoing authoring, storage, enrichment, publishing, consumption and maintenance of master data. It specifies the processes, roles, standards and metrics that ensure the effective and efficient use of master data in enabling an organization to achieve its goals. When choosing the members to MDM governance group, it is good to choose employees from different parts of the organization to ensure a wider view on the matter.

4. **Organization:** A unified MDM program will create change. Therefore, communication, training and change management will become challenges and they must be planned and resourced with appropriately skilled people. The MDM program will require a matrix organizational approach, with leadership coming from the business side and the IT organization being a major contributor. At the core of the MDM program there should be a small number of dedicated resources in a central team, the so called steering group. The appointment of data stewards, who have responsibility for the quality of master data, will be the key to success [7].

5. **Technology Infrastructure**: In this step it needs to be determined what technologies are needed to enable the MDM vision and strategy, and where and how this technology should be sourced.

6. **Metrics:** In the Gartner research it is said that without measuring the quality of master data and its effects on business performance, there is no objective basis for reporting improvements. The top level of an organization's performance management metrics are the corporate goals, which are typically revenue, profit and market share related. On the next layer are strategic metrics related to the operational effectiveness, customer intimacy, and product or service

leadership goals. On the next layer there are process-level metrics. MDM improvements should have a positive, direct effect on the majority of these metrics.[8]

The objectives of Master data Management is to improve business processes by improving the quality of data. Errors can be reduced in order process by improving Master data quality.
The following parameters are evaluated to enrich quality of data. [1]

- Accuracy
- Completeness
- Update Status
- Relevance
- Consistency
- Source
- Accessibility

MDM is a strategy that provides a single, unified, and trusted view of master data entities for any an application. In other words, MDM is a set of software solutions that manage the creation, governance, delivery, and use of master data across the organization [10].

Master data management (MDM) is a branch within the field of Enterprise Information management (EIM) focusing on master data. It is comprised by a combination of processes, standards, governance, organization and technologies with the objective of providing consumers of master data across the enterprise with a single – authoritative - view of its master data. Many times an organizations master data is scattered across a number of disparate data sources. Where master data consist of those key information objects for which maintaining good quality is critical as they form the basis for business processes. [9].

.1.2 Process of Master Data Creation

After an organization identifies various entries for similar data, it must be integrated to a link and duplicates must be resolved to ensure that processes are not creating duplicates. In this pattern, customer on boarding, account creation, vendor on boarding, and other typical business

processes push the master data that they author into MDM and search MDM to update existing master data. By applying match and merge technique and evaluating trust factor one can make a golden copy of record.

The process of master data creation depending on the consumer can be categories in to one of three key patterns of use [10].

- **Collaborative Authoring**. As authoring of a master data object may can be conducted by several people, this gives rise to a highly complex workflow, in where a multitude of topics must be agreed on by a group of people. Typical example is that of product information management (PIM) systems. Due to the complexity in developing and managing products, requiring several parties to agree, PIM systems commonly support the collaborative style. Furthermore, for a collaborative MDM style to executeeffectively this requires core capabilities such as task management, and state management to guide and monitor tasks being collaborated on. Since task may be worked on concurrently there is a need for control mechanisms to preserve data integrity.

- **Operational MDM** It is a solution intended to manage transactional data used by operational applications. This is realized by providing stateless services that can be invoked by business processes or directly by applications. The nature of the operational MDM thus makes it suitable to be incorporated into a service-oriented architecture environment [6].

- **Analytical MDM**. As sound and strategically good descriptions are of paramount importance for any organization, analytical MDM has surfaced as means of managing the enterprise's analytical data. The goal of analytical MDM provides high quality dimensions with their multiple simultaneous hierarchies to data warehousing and BI technologies. As further explained analytical MDM is essentially about the intersection with Business Intelligence (BI) of which three such intersections are

  (i) Trusted data source

  (ii) Analytics on MDM data

(iii) Analytics as a key function in a MDM system [6].

For BI tools to be able to provide meaningful analysis the data on which these tools operate must be good and trustworthy. Furthermore, as analytical MDM systems provide to means of improving the data quality in a heterogeneous application landscape these shortcomings in data quality will also make its way in to the BI tools.

1.3 Master Data Governance

Master data governance is the decision and action processes of managing shared resource for common good, which includes:

- Identifying shared resources that would benefit from being managed.
- Communicate who is responsible for a resource; how the governing bodies are organized and the processes they will follow.
- Distributing decision making with regards to what is encouraged and likewise discouraged when touching upon shared resources
- Assigning roles and responsibilities to different parts of an implementation involving processes and policies for a shared resource [5].

Creating master data and using it within business processes are the two primary points where MDM is used within a business. Data Governance means assurances that data is a trusted asset that is ready for use within their business processes. By using master data governance process in order to enforce these policies there is assurances to business owners that master data is not only accurate but also supports their usage requirements. Without these polices, consumers lose confidence that the data complies with their usage requirements and lose trust in master data. The objective of Master data governance process is to enforce certain policies for effectiveness of result [8].

2. Research Methodology

A research is about establishing facts and reaching new conclusions by systematic study of a system. Research is conducted by taking some objectives in mind. In research an attempt is made

to answer question(s). This gives an overview of these research methodologies. The types of methodologies used in the research depend upon on the research itself. After deciding methodology comes collection of data. Data collected can be primary data that is raw data not used before and it can be secondary data. For conducting a research tool plays a very handy role for simulating research finding into real time environment. Section 2.1 states the problem under study. Section 2.2 outlines the objectives of the study. Section 2.3 gives the scope of the study.

2.1 Problem Statement

Presented problem is entitled as "A Novel approach for the validation of proposed Database Using Master Data Management. In any organization, the interface for collecting the customer data may be phone, email, over the counter, form etc. The information collected is the basis for the high-value, core information that is used to support critical business processes across the enterprise. Regardless of how good the quality of the information is that an organization maintains, the desired goals cannot be realized if the main business processes are inefficient, do not meet client needs and difficult to adapt to meet new business challenges. Thus the ultimate solution is to adopt MDM strategies to resolve the problem of inconsistent data. MDM is a discipline that provides a single, unified, and trusted view of master data entities for any user or application.

2.2 Objectives of Research

Research Objectives declares clearly the aim of research and its purpose. The main objectives are:

1. To study existing Master data management techniques and to know the loopholes there in.
2. To identify data quality determinants for the enrichment of database.
3. To propose a technique for creating Master database.
4. To design data monitoring policies for validation of Master database.
5. To propose a technique for Master data consumption.
6. To compare the proposed technique with select one.

2.3 Scope of the study

The study deals the management, maintaining and validating practices of larger database. This research includes various strategies for improving data quality and consistency. By maintaining a golden copy of record using MDM, the business and IT organization work together to ensure the uniformity, accuracy, semantic persistence, stewardship and accountability of enterprise's official, shared master data. Organizations can apply MDM to eliminate the costly debates on "which data is right" which can lead to effective decision making and business performance.

2.4 Research Methodology Used in Research

In the methodology section, the research approach and research process chosen for the thesis is discussed. Additionally, the method used for the selection of literature and empirical objects, and the validity and generalizability of the thesis are described. The thesis is based on two studies: literature study and empirical study. The literature study covers literature regarding Customer Data Management, Master Data Management, Data Governance, Data Stewardship, Data Quality Management and Data Quality Assessment and Improvement. The empirical study is based on a case company.

Initially, aim for the thesis is defined through collaboration with the thesis assignor and using existing literature. An initial literature study was done to understand the concept of MDM and Data Quality Management. The essential factors found in the initial literature study would be the base for the development of empirical study. The literature study is then followed by an empirical study based on a case company. The findings would be calculated and finally it would be compared with older techniques.

2.5 Chapterization

Chapter 1: Introduction

This chapter will be presenting an introduction of the master data management and its policies. It covers the basic factors for determining the qualities of database, how to improve that and the process of validation. The parameters on the basis of which master data is managed. At the end an insight into application areas of Master data management shall be given.

Chapter 2: Review of literature

This chapter will be presenting the literature review relevant to the present research work directly or indirectly from various sources such as reference books, journals, periodicals, research oriented user groups and internet.

Chapter 3: Research Methodology

This chapter will comprise of problem statement, research objectives, scope and significance of the study, research design, analysis, chapterization and limitations of the study.

1. To identify data quality determinants for the enrichment of database.
2. To propose a technique for creating Master database.
3. To design data monitoring policies for validation of Master database.
4. To propose a technique for Master data consumption.

Chapter 4: Data quality Determinants for the Enrichment of Database.

This chapter will discuss the set of determinants which are useful for efficient storage of data. Various techniques will be discussed in order to enrich the database by improving the quality of entered data.

Chapter 5: Process of Master Data Creation

This chapter will be presenting the technique to create a master data of a record among multiple entries. On the basis of match – merge, a trust value is evaluated in order to create a golden copy of record.

Chapter 6: Designing of Data Monitoring Policies for Validation of Master database

This chapter will be focusing on the designing of a monitoring polices in order to check the validity of created master data.

Chapter 7: Technique for Master data Consumption

This chapter will be focusing on the designing of a technique for the consumption of master database.

Chapter 8: Comparison of Proposed technique with select one

This chapter will be comparing the proposed technique of creating and validating master data with selective one.

Chapter 9: Conclusion and future scope

Finally, this chapter will be devoted to the concluding remarks and future prospects of the present work.

3. LITERATURE REVIEW

It presents the review of work done by various researchers that provides a useful insight for managing the master data management. This work done is very helpful in understanding what has been done in the field of validating master data and how it can be further improved.

Andy Papa (2017) stated that the scope of MDM is big topic for global enterprises whose local flexibility plays a direct role, therefore it is considered a big change inside the organizations and sometime the interviewees are a bit reserved and skeptical towards it. The purpose of this research is to analyze the different components that affect MDM in global organizations. The research has been developed on a global financial organization where the master data concept is vague but its importance is easily visible. The three main categories discovered are relevant to global organization and therefore they can be a good starting point also for other different organization. These three main categories could be divided into specific other themes and be compared to the eleven ones discovered in this research in accordance with the benefits of MDM. [1]

Amber Lee Dennis (2017) discovered that the implementation of “Stibo Systems” STEP(MDM platform) solution for Master Data Management (MDM), includes 60% reduction in new product time-to-market, and improved efficiency because data only needs to be input once, regardless of the source. STEP is a Master Data Management (MDM) platform which integrates multiple disparate systems by integrating the process of aggregating and consolidating information around an organization’s products, customers, suppliers, employees, assets, location and reference data from multiple sources and formats. STEP is the easy way to connect

information sources in order to derive actionable insights, and publishes it to backend systems as well as online and offline channels. [2]

Jochen Kokemüller et. al. (2015)discovered the characteristics of different organizational data. They categorize the data present in enterprise, namely transactional and inventory data. This is the domain of the discipline of master data management with a discussion of architectures for master data management, before one can analyze commercial solutions. Two results are presented which are basically surveys that is executed with the system vendors. Discriminating features could be found in the field of data modeling and target industries. Additionally, it is proved that supported data exchange formats are very diverse which describe that the vendors follow different strategies. While the purpose of master data management aims is to raise and maintain high information quality which the systems often provide out-of-the-box only for basic features for this task. Yet, most systems may work together with specialized solutions in this domain. [3]

Jones Eyob (2015) discussed the importance of MDM primarily on the governance and organizational dimensions. As per the result, the most common problem for failing of properly managing these dimension according to literature. A study on NCC revealed that previous attempts embarking on MDM have not been able to successfully follow-through, in part, due to inability to generate a compelling business case and agree on a target data model. From the challenges and proposed success factor one can also conclude that a majority of the challenges are related to the absence of a data governance function focusing on master data as well as appointed roles responsible for the different master data entities. For this reason in this research a reference model for the governance organization was proposed encompassing specific roles. [4]

Pavan Kumar Purohit (2014) concluded that MDM cannot be classified as only an IT problem but it is a managerial challenge which requires structural changes to manage business processes, and managerial decision making. Without engagement of the business owners MDM cannot be succeed. In the case study it was found that the funding agency can now accurately track the number clients that were served. Thus MDM improves overall data quality and confidence. With success implementation of MDM, a single source of truth exists regarding any master data that is tracked – customer, products, suppliers etc. This will help businesses make proper business decisions with the confidence that the data on which these decisions are based is accurate. [5]

Katri Vihavainen (2014) gave an idea to provide guidelines and best practices on how to achieve and maintain high level integrity of customer data. In this thesis, different aspects of Master Data Management and data quality are studied, followed by a current state analysis, and then practical guidelines are provided based on the existing literature. According to the research, data cleaning actions needed to be reported systematically and as accurately as possible. This must be done in a way that the actions done are clearly stated so that a person outside the cleaning project also understands them. They stated that data cleaning system is used, it needs to be able to not only screen and filter the data against data quality and business rules, but also make suggestions or even decisions based on very detailed rules and additionally validating the data against external data bases. [6]

Delia Lucas (2014)stated that 76% of the problems related to poor Data Quality are caused by incorrect data entries. In this research Master Data Management with different approaches and solutions were discussed, but main focus was on the technical scope. As per the results, for a good database ERP structure is necessary in order to "attack" the problem, but if according to the surveys the main cause of poor Master Data is simply data entries, maybe the problem is not well framed. Thus one should take proper care should while entering the data. [7]

Gholam Ali(2014) concluded that CDM (customer data management) is the best option with respect to factors such as cost and availability. The main focus was specifically on gathering information on fundamental key factors in implementation of an Data Management; the key factors are cost, schedule, performance, efficiency, limitations, risk, training, operations, compliances, security, accessibility, dependability, data quality, deployment, stability, maintainability, reliability, availability, flexibility, scalability, and predictability metrics. However, FDM may be considered as short term solution. To reduce the cost at the beginning, FDM should be used as a short-term practice to drive quick benefits to the organization before moving to longer-term strategy – CDM. [8]

Tapan kumar Das (2011) discussed the challenges related to Master data. From the study it is concluded that the main challenges are because of presence of duplicate CDIs, data and application integration, data consolidation, data governance challenges, metadata management. As MDM is about implementing a technology, also the problems were not only technology based, business initiatives is needed in the MDM arena. Clean, consolidated and accurate master

dataset aimlessly propagated throughout the enterprise can save millions of dollars, increase market base, improve customer loyalty and support sound corporate governance. [9]

Half Anude [2011]stated that the Customer Master Data team of the case company may use the results in their data cleaning projects. In case if employee's changes, or information about the previous projects is not effectively reported, project team can find assistance from the guidelines provided by this research. Secondly, other users of master data can use this guide. There are different teams handling master data in the case company, for example Vendor Master Data team, APAC and NA regional Customer Master Data teams, and business specific master data teams both in the core system and business target systems. In addition, the research results can be used outside the case company when adjusted to their own master data system and business processes. [10]

Strout, Eisenhauer [2011] developed a Master Data Management project for a common vocabulary with appropriate policies and rules in order for the business to keep data organized and consistent. Data needs to be consistent from one person to another, consistent over time, consistent from system to system, and from one department to another. In the case company, there has not been common processes for data management with the consequence that multiple systems and redundant processes exist that can lead to inconsistencies. "Business Agility" includes the ability of an organization to enter new markets or current markets with new products as quickly as necessary. MDM also helps in the creation of a single version of the "truth", and managing the quality of the data that makes up that "truth" is a significant issue for almost any company. [11].

Walker et. al. (2009) defined master data as being the core data of a company; customers, vendors, products and employees. Whereas Master data management is a combination of 7 data governance, business processes, data quality, data enrichment, and a technical solution. It is a companywide program that requires co-operation of a company's business units and business areas. MDM also must ensure the availability of accurate and up-to-date master data that is of good and trusted quality. [12]

4. REFERENCES

[1] AndyPapa,"Master Data Management in Global Enterprise" May, 2017.

[2] Amber Lee Dennis, “Enhanced Master Data Management Implemented at Office Depot” August 31, 2017.

[3] Jochen Kokemüller, Fraunhofer, Anette Weisbecker “Master Data Management: Products and Research.” 2015.

[4] Jonas Eyob, “Master data management: A study of challenges and success factors at NCC”. 2015.

[5] Pavan Kumar Purohit, “Master Data Management (MDM) – Strategies, Architecture and Synchronisation Techniques, March 31, 2014.

[6] KatriVihavainen, “Managing Data Integrity as Part of Master Data Management” 2014.

[7] Delia Rodríguez Lucas,” Master data management as a tool to improve business Profitability, 2014.

[8] Gholam Ali Shaykhian, Mohd Abdelgadir Khairi,“Factors influence data management models selection, 2014.

[9] Tapan kumar Das and Manas Ranjan Mishra, “A Study on Challenges and Opportunities in Master Data Management” in International Journal of Database Management Systems, Vol.3, No.2, May 2011.

[10] Half AnudeScheidi ,”Master data management Maturity and Technology Assessment” 2011.

[11] Strout, Steven B. and Eisenhauer, John A.,” A Project for master data management” 2011.

[12] Walker, Andy and Ganapathy, Jagadeeshwaran,“Effective Master Data Management with SAP , 2009.

[13] Schönemann, D, "P2P Architecture for Ubiquitous Supply Chain Systems", 17th European Conference on Information Systems (ECIS), Verona, Italy, 2009

[14] Kokemüller, Weisbecker, A. "A Mobile Support System for Collaborative Multi-Vendor Sales Processes", Proceedings of the Fourteenth Americas Conference on Information Systems, Toronto, ON, Canada, August 14th-17th, 2008 .

[15] Walter, P, Werth, D. Loos, "Peer-to-Peer-Based Model-Management for Cross-Organizational Business Processes", Los Alamitos, CA, USA, June, 2006.

[16] Chen, Z., Kalashnikov, D. V. Mehrotra, S. "Exploiting relationships for object consolidation", Proceedings of the 2nd international workshop on Information quality in information systems, New York, NY, USA, pp. 47-58, 2005.

[17] Flint, D. “How Nestle Reduced Errors in its Management of Master Data”, Gartner Group Research.,2004.

[18] Gephart, R. P. “Qualitative research and the Academy of Management Journal”, Academy of Management Journal, Vol. 47,2004.

[19] Erasala, N., Yen, D. C. Rajkumar, T. M. "Enterprise Application Integration in the electronic commerce world", Computer Standards und Interfaces, 25, 2, pp. 69 – 82, 2003.

[20] Eschinger, C., “Report Highlight for Market Trends: Master Data Management Growing Despite Worldwide Economic Gloom,” Gartner, 2003.

[21] Schönemann, N., Fischbach, K. Schoder, D. "P2P Architecture for Ubiquitous Supply Chain Systems", 17th European Conference on Information Systems (ECIS), Verona, Italy, 2003.

[22] Walter, P., Werth, D. Loos, P. "Peer-to-Peer-Based Model-Management for Cross-Organizational Business Processes", Los Alamitos, CA, USA, pp. 255-260,2003.